Supplementary Materials

Common Genetic Variants Modulate Pathogen-Sensing Responses in Human Dendritic Cells

Mark N. Lee, Chun Ye, Alexandra-Chloé Villani, Towfique Raj, Weibo Li, Thomas M. Eisenhaure, Selina H. Imboywa, Portia I. Chipendo, F. Ann Ran, Kamil Slowikowski, Lucas D. Ward, Khadir Raddassi, Cristin McCabe, Michelle H. Lee, Irene Y. Frohlich, David A. Hafler, Manolis Kellis, Soumya Raychaudhuri, Feng Zhang, Barbara E. Stranger, Christophe O. Benoist, Philip L. De Jager, Aviv Regev*, Nir Hacohen*

*Corresponding authors. E-mail: aregev@broadinstitute.org (A.R.); nhacohen@partners.org (N.H.)

This PDF file includes:

Materials and Methods
Supplementary Text
Figs. S1 to S5
Captions for tables S1 to S11
References *(46-72)*

Other Supplementary Materials for this manuscript includes the following:

Tables S1 to S11 as zipped archive:

- Table S1. Samples.
- Table S2. Differential expression analysis.
- Table S3. Gene signature set.
- Table S4. cis-eQTLs and cis-reQTLs from pooled analysis.
- Table S5. Imputation meta-analysis.
- Table S6. Conditioning.
- Table S7. ChIP-Seq enrichment.
- Table S8. Trans-eQTLs and trans-reQTLs from pooled analysis.
- Table S9. Trans-eQTLs and trans-reQTLs from pooled analysis (only cis considered).
- Table S10. rs12805435 trans.
- Table S11. GWAS.

Materials and Methods

Study subjects

Donors were recruited from the Boston community as part of the Phenogenetic Project and ImmVar Consortium, and gave written informed consent for the studies. Individuals were excluded if they had a history of inflammatory disease, autoimmune disease, chronic metabolic disorders or chronic infectious disorders. For the microarray study, 30 healthy human donors were recruited. Donors were between 19 and 49 years of age; 15 were male and 15 were female; 18 were Caucasian, 6 were African-American and 6 were Asian; 12 provided a serial replicate blood sample (i.e. separate blood draw 2-9 months after the first collection). For the Nanostring study, 534 healthy donors were recruited, 37 of whom came in multiple times providing additional serial replicates (i.e. separate blood draw > 1 month later). Donors were between 18 and 56 years of age (avg. 29.9); demographics are listed in table S1.

Reagents

LPS (ultrapure lipopolysaccharide from *E. coli* K12) was obtained from Invivogen (San Diego, CA). Influenza A (PR8 ΔNS1) was prepared as described (*46*). Recombinant human IFNβ was obtained from PBL Assay Science (Piscataway, NJ). Antibodies used were anti-IRF1 (sc-497x; Santa Cruz Biotechnology; Dallas, TX), anti-STAT2 (sc-476x; Santa Cruz Biotechnology) and anti-IRF9 (sc-10793x; Santa Cruz Biotechnology).

Preparation and stimulation of primary human monocyte-derived dendritic cells

35-50 mL of peripheral blood from fasting subjects was collected between 7:30-8:30 am. The blood was drawn into sodium heparin tubes and peripheral blood mononuclear cells (PBMCs) were isolated by Ficoll-Paque (GE Healthcare Life Sciences; Uppsala, Sweden) centrifugation. PBMCs were frozen in liquid N_2 in 90% FBS (Sigma-Aldrich; St. Louis, MO) and 10% DMSO (Sigma-Aldrich). Monocytes were isolated from PBMCs by negative selection using the Dynabeads Untouched Human Monocytes kit (Life Technologies; Carlsbad, CA) modified to increase throughput and optimize recovery and purity of $CD14^{+}CD16^{lo}$ monocytes: the FcR Blocking Reagent was replaced with Miltenyi FcR Blocking Reagent (Miltenyi; Bergisch Gladbach, Germany); per mL of Antibody Mix, an additional 333 ug biotinylated anti-CD16 (3G8), 167 ug biotinylated anti-CD3 (SK7) and 167 ug biotinylated anti-CD19 (HIB19) antibodies (Biolegend; San Diego, CA) were added; the antibody labeling was modified to be performed in 96-well plates; and Miltenyi MS Columns or Multi-96 Columns (Miltenyi) were used to separate magnetically-labeled cells from unlabeled cells in an OctoMACS Separator or MultiMACS M96 Separator (Miltenyi) respectively. The number of PBMCs and monocytes was estimated using CellTiter-Glo Luminescent Cell Viability Assay (Promega; Madison, WI). A subset of the isolated monocytes was stained with PE-labeled anti-CD14 (M5E2; BD Biosciences; Franklin Lakes, NJ) and FITC-labeled anti-CD16 (3G8; Biolegend), and subjected to flow cytometry analysis using an Accuri C6 Flow Cytometer (BD Biosciences). A median of 94% $CD14^{+}$ cells and 99% $CD16^{lo}$ cells was obtained. The remaining monocytes were cultured for seven days in RPMI (Life Technologies) supplemented with 10% FBS, 100 ng/mL GM-CSF (R&D Systems; Minneapolis, MN) and 40 ng/mL IL-4 (R&D Systems) to differentiate the monocytes

into monocyte-derived dendritic cells (MoDCs). $4x10^4$ MoDCs were seeded in each well of a 96-well plate, and stimulated with 15 ng/mL LPS for 5 h, influenza virus for 10 h, 100 U/mL IFNβ for 6.5 h or left unstimulated for up to 4 conditions per donor depending on available cell number. Cells were then lysed in RLT buffer (Qiagen; Hilden, Germany) supplemented with 1% β-mercaptoethanol (Sigma-Aldrich).

Microarray gene expression profiling

Total RNA was isolated from MoDCs using the RNeasy Mini kit (Qiagen). RNA quantification and quality were assessed using Agilent 2100 Bioanalyzer (Agilent Technologies; Santa Clara, CA). The cDNA synthesis, labeling and subsequent hybridization to the microarrays were performed by the company Expression Analysis (Durham, NC). Affymetrix Human Gene 1.0 ST arrays (Affymetrix; Santa Clara, CA) were used to obtain genome-wide expression profiles.

Normalization of microarray data, adjustment for confounding effects and differential expression analysis

Outlier detection was first performed. The dataset was then normalized using quantile normalization as part of the RMA pipeline. Expression > $\log_2(80)$ was used as a filter resulting in 16,514 of 26,982 annotated transcripts. To eliminate possible confounding effects – in particular the batch effect from running arrays over two plates – surrogate variable analysis (SVA) was used to identify 13 permutation-significant surrogate variables (SVs) while controlling for known covariates for stimulation, population, gender and race. These SVs were used for subsequent analysis. K-means clustering was used to group the SVA-adjusted data into four distinct clusters corresponding to genes up-regulated in one or both of the conditions and a set of genes that were down-regulated. The number of differentially expressed transcripts per condition was computed using an Empirical Bayes model that included the SVs as covariates. Significance of differential expression was assessed by FDR (0.01) and $\log_2$(fold change) restrictions (> 0.75 or < –1.5). Bonferroni adjusted binomial *P*-values for over-representation of Gene Ontology pathways were reported.

Mixed model estimate of inter-individual and inter-population variable genes

A linear mixed model was used to model the variance components for the expression of a gene *i* in individual *j* as:

$$y_{ij} = \beta_0 + \beta_1 \text{Sex} + \beta_2 \text{Age} + \beta_3 \text{Population} + \sum_{k=1}^{n} \rho_k PC_k + u_j + \epsilon_{ij}$$

where $Var(\boldsymbol{u}) \sim \sigma^2_{inter} K$ and $Var(\boldsymbol{\epsilon}) \sim \sigma^2_{\epsilon} I$, K is a $N \times N$ matrix whose entry is 1 if two samples are the same individual. The above model was tested using a likelihood ratio test for inter-individual variability ($\sigma^2_{inter} = 0$) as well as population specific genes ($\beta_3 = 0$). Each condition (baseline, LPS, FLU, IFNβ and the respective responses) was analyzed separately. For each test, permutation based *P*-values were computed by randomly shuffling the individual labels 200 times and recalculating the likelihood ratio statistic. FDRs were computed based on the empirical *P*-value distributions.

Selection of gene signature

Normalization controls were selected by choosing genes with low variance in expression in the microarray data. Additional low variance controls were selected by choosing genes with low variance in response to LPS or influenza stimulations. Genes not known to be expressed in MoDCs including *CD3*, *CD19*, *CD56* and *CD235a* were added as controls for contaminated cell types. Sex-specific genes from the Y-chromosome including *DDX3Y*, *EIF1AY* and *ZFY* were added for sample QC. Positive controls that represented genes with known *cis*-eQTLs were selected from baseline monocyte *cis*-eQTL data (*47*) and baseline MoDC *cis*-eQTL data (*48*). Known pathway components were selected from Ingenuity Pathway Analysis and references including (*49-52*). To capture inter-individual variation, genes with a $\log_2$(fold change) > 0.75 or $\log_2$(fold change) < –1.5, and with an inter- vs. intra-individual FDR < 0.1 in (i) LPS-induced expression values, (ii) influenza-induced expression values, (iii) LPS response values and (iv) influenza response values were selected. To capture variation in population, genes with a $\log_2$(fold change) greater than 0.5 or less than –1.5, and with an inter- vs. intra-population FDR < 0.1 in LPS response values and influenza response values were selected. Genes with a maximum value of $\log_2$ > 13 across the microarray dataset were excluded to prevent saturation. Probes that excluded common SNPs (MAF > 5% from 1000 Genomes) were chosen, except for 11 genes for which this was not possible (table S3). The ability of the signature gene set to predict the genome-wide expression profile was measured using a structured regression model that imposes sparsity constraints as implemented in the PSI (*53*) package for expression imputation.

Nanostring gene expression profiling and data normalization

The Nanostring nCounter system (Nanostring; Seattle, WA) was used to digitally count transcripts in a multiplex reaction as previously described (*54*). Lysates in RLT buffer were hybridized for 12-24 hours with custom nCounter Gene Expression CodeSets. Quantification of hybridized RNA was performed using the nCounter Analysis System. To normalize the nCounter data, each sample was first normalized using the internal positive spike-in controls; the data was then normalized based on the average expression level of the low variance control genes. The per sample dispersion parameter was estimated as previously described (*55*) and the data was variance-stabilized for further processing.

DNA extraction and genotyping

Genomic DNA was extracted from 5 mL whole blood (DNeasy Blood & Tissue Kit; Qiagen), and quantified by NanoDrop (Thermo Fisher Scientific; Waltham, MA). Each subject was genotyped using the Illumina Infinium Human OmniExpress and Human Exome BeadChips (Illumina; San Diego, CA), which includes genome-wide genotype data as well as genotypes for rare variants from 12,000 exomes as well as common coding variants from the whole genome. In total, 951,117 SNPs were genotyped, of which 704,808 SNPs are common variants (Minor Allele Frequency (MAF) > 0.01) and 246,229 are part of the exomes. The genotype success rate was greater than or equal to 97%. We applied rigorous subject and SNP quality control (QC) that includes (1) gender mis-identification, (2) subject relatedness, (3) Hardy-Weinberg Equilibrium testing, (4) use concordance to infer SNP quality, (5) genotype call rate, (6) heterozygosity outlier

and (7) subject mismatches. In the European population, we excluded 1,987 SNPs with a call rate < 95%, 459 SNPs with Hardy-Weinberg equilibrium *P*-value < 10^{-6}, 234 SNPs with a MisHap *P*-value < 10^{-9}, and 63,781 SNPs with MAF < 1% (a total of 66,461 SNPs excluded). In the African-American population, we excluded 2,161 SNPs with a call rate < 95%, 298 SNPs with Hardy-Weinberg equilibrium *P*-value < 10^{-6}, 50 SNPs with a MisHap *P*-value < 10^{-9}, and 17,927 SNPs with MAF < 1% (a total of 20,436 SNPs excluded). In the East Asian population, we excluded 1,831 SNPs with a call rate < 95%, 213 SNPs with Hardy-Weinberg equilibrium *P*-value < 10^{-6}, 47 SNPs with a MisHap *P*-value < 10^{-9}, and 84,973 SNPs with MAF < 1% (a total of 87,064 SNPs excluded). After QC, 52 subjects across all three populations and approximately 18,000-88,000 SNPs in each population were filtered out from our analysis.

Population stratification: Underlying genetic stratification in the population was assessed by multi-dimensional scaling using data from the International HapMap Project (CEU, YRI and CHB samples) combined with IBS cluster analysis using the Eigenstrat 3.0 software (*56*).

The quality control of the genotyping data were performed using PLINK (*57*).

Pooled condition-specific eQTL mapping

Using only typed SNPs, we performed eQTL mapping by pooling all individuals across three populations together for each of the four conditions (baseline, LPS, FLU and IFNβ) and the corresponding fold change (LPS/baseline, FLU/baseline and IFNβ/baseline). We included gender, age and population as known covariates and significant principal components (determined by permutation) to account for unknown confounding effects in the following linear model:

$$y_{ij} = \beta_0 + \beta_1 \mathrm{SNP} + \beta_2 \mathrm{Sex} + \beta_3 \mathrm{Age} + \beta_4 \mathrm{Pop} + \sum_{k=1}^{n} \rho_k PC_k + \epsilon_{ij}$$

We estimated the parameter β_1 and its corresponding standard error for each gene-SNP pair in each condition. We computed a likelihood ratio statistic comparing β_1 to the null hypothesis of $\beta_1 = 0$. All mapping were performed using a modified version of Matrix eQTL (*58*).

For *cis*-eQTLs, we computed experiment-wise significance for the best eQTL detected per gene by permuting the labels of individuals 200 times and computing the FDR by comparing the distribution of *P*-values to the empirical null using the qvalue package (*59*). We used random effects meta-analysis that accounts for effects size heterogeneity between conditions and the M-value (*60*) to categorize genes into specific patterns of association. A *cis*-eQTL is categorized to be specific to a particular condition if its M-value in the condition is > 0.9. For *cis*-reQTL, a conservative exclusion criteria of M-value < 0.1 was additionally used. In addition, a reQTL was required in one condition.

For *trans*-eQTLs, we considered all variants as well as only *cis*-variants to genes on our codeset to limit the number of multiple tests performed. To detect additional independent signals, we conditioned on the top five *cis*-eQTLs by including them as additional

covariates in the regression model. For all of these tests, we assessed experiment-wise significance for the best *trans*-eQTL detected or best secondary *cis*-eQTL per gene by permuting the labels of individuals 200 times and computing the FDR by comparing the distribution of *P*-values to the empirical null using the qvalue package.

Genotype imputation

To accurately evaluate the evidence of association signal at variants that are not directly genotyped, we used the BEAGLE software (version: 3.3.2) (*61*) to impute the post-QC genotyped markers using reference Haplotype panels from the 1000 Genomes Project (The 1000 Genomes Project Consortium Phase I Integrated Release Version 3), which contain a total of 37.9 million SNPs in 1,092 individuals with ancestry from West Africa, East Asia and Europe. For subjects of European and East Asian ancestry, we used haplotypes from Utah residents (CEPH) with Northern and Western European ancestry (CEU), and combined panels from Han Chinese in Beijing (CHB) and Japanese in Tokyo (JPT), respectively. For imputing African-American subjects, we used a combined haplotype reference panels consisting of CEU and Yoruba in Ibadan, Nigeria (YRI). After genotype imputation, we filtered out poorly imputed (BEAGLE $r^2 < 0.1$) and low MAF SNPs (MAF < 0.01), which resulted in 7.7M, 6.6M, 12.7M common variants in European, East Asian and African-American, respectively. This set of genotyped and imputed markers was used for all the subsequent association analysis.

Population-specific eQTL mapping using imputed SNPs for fine-mapping

To increase the power to detect functional variants, we performed fine-mapping using imputed SNPs in each population (Asian, African-American and Caucasian) and condition (baseline, LPS, FLU, IFNβ, LPS/baseline, FLU/baseline and IFNβ/baseline) separately using the following linear model that includes significant principal components estimated in each population/condition separately:

$$y_{ij} = \beta_0 + \beta_1 \text{SNP} + \beta_2 \text{Sex} + \beta_3 \text{Age} + \sum_{k=1}^{n} \rho_k PC_k + \epsilon_{ij}$$

We estimated the parameter β_1 and its corresponding standard error for each gene-SNP pair in each population/condition. For each condition, we used mixed effects meta-analysis (*60*) that accounts for differing linkage disequilibrium patterns in each population (based on dense imputed data from 10M SNPs using the 1,000 Genomes Project dataset (*62*)) and the assumption of shared causal alleles (*63*) (using a random effects meta-analysis (*60*) on the imputed *cis*-reQTLs from the three populations). Next, we searched for associated variants (at each locus) that alter TF binding elements in these regions to detect variants with shared effects across the three populations, hypothesizing that these might be enriched for functional variants in the absence of epistasis.

ChIP-Seq enrichment

SNPs intersected with human ChIP-seq peaks from ENCODE (*64*) were obtained from HaploReg (*65*). Mouse ChIP-seq data were obtained from (*66*) and the liftOver tool was used (with parameters -minMatch=0.1 and -multiple) on the mouse peaks to find orthologous regions between mouse genome version mm9 and human genome version hg19. Lifted over peaks were then intersected with human SNPs using BEDTools (*67*).

For each ChIP annotation, the overlap with the top five eQTLs for each gene was compared to the frequency of the ChIP annotation among a background set of all 1000 Genomes Project Phase 1 SNPs (*68*) with frequency greater than 5% in at least one population within 1 Mb of the start and end of assayed genes. Enrichment relative to background was evaluated using a binomial test.

Allelic imbalance

Total RNA was extracted using the RNeasy Mini kit (Qiagen) from MoDC lysates from individuals heterozygous for selected exonic SNPs as well as for the most significant *cis*-eQTL SNP in LD. cDNA was synthesized using High Capacity cDNA Reverse Transcription kit (Life Technologies). TaqMan genotyping assays (Life Technologies) that discriminate between the two exonic alleles were designed and performed on the cDNA and corresponding genomic DNA (gDNA) samples using the LightCycler 480 system (Roche; Basel, Switzerland). Ratio of alleles in the cDNA was normalized to the ratio of alleles in the gDNA. Amplification primers for *SLFN5*: 5'-CGTTTTCTCTTGCGGAATGGT-3' and 5'-TCATTGCTGTTAGAAGCCTGTCTTT-3'; reporter primers for *SLFN5*: 5'-CAATATCCTTCGGAGAATA-3' (VIC) and 5'-CAATATCCTTCAGAGAATA-3' (FAM). Amplification primers for *CLEC4F*: 5'-CATCACCACTTTGGCAGGGA-3' and 5'-CCATGTGGCCTTTAAATGTCTGGAT-3'; reporter primers for *CLEC4F*: 5'-AGAAATGCGAGAGCTT-3' (VIC) and 5'-AGAAATGCAAGAGCTT-3' (FAM).

cDNA overexpression

Constructs encoding human *IRF7* or *eGFP* control cDNA were nucleofected into MoDCs using Amaxa Human Dendritic Cell Nucleofector Kit (Lonza; Basel, Switzerland). 18 h later, cells were infected with influenza and then lysed 10 h post-infection in RLT buffer (Qiagen) supplemented with 1% β-mercaptoethanol. Constructs encoding human *IRF7* or *eGFP* control cDNA were also transfected into HEK-293 cells using *Trans*IT-LT1 (Mirus; Madison, WI). Cells were lysed 22 h later in RLT buffer supplemented with 1% β-mercaptoethanol.

Luciferase reporter assays

The major haplotypes of *SLFN5* (192 bp; chr17:33,571,450-33,571,641), *CLEC4F* (159 bp; chr2:71,050,488-71,050,646) and *ARL5B* (199 bp; chr10:18,947,302-18,947,500) were cloned 5' of the minimal promoter in the firefly luciferase reporter vector pGL4.23 (Promega). Minor allele SNPs were introduced using the QuikChange Lightning Site-Directed Mutagenesis Kit (Agilent Technologies). Constructs were co-transfected with the *Renilla* luciferase control vector pRL-CMV (Promega) into HEK293T cells using *Trans*IT-LT1 (Mirus). 8 h later, cells were stimulated with 1000 U/mL recombinant IFNβ for 21 h, and then firefly and *Renilla* luciferase activities were measured using Dual-Glo Luciferase Assay System (Promega).

Gel shift assays

Nuclear extracts were prepared (NE-PER; Thermo Fisher Scientific) from MoDCs stimulated for various times with 100 U/mL IFNβ. 25-27 nt DNA oligos (IDT; Coralville,

IA) were labeled with [γ^{32} P]-ATP using T4 polynucleotide kinase (NEB; Ipswich, MA), and annealed with the reverse complementary oligos. Sequences used were as follows: ISRE control (5'-AAGTACTTTCAGTTTCATATTACTCTA-3'; Santa Cruz Biotechnology), ISRE mutant (5'-AAGTACTTTCAGTGGTCTATTACTCTA-3'; Santa Cruz Biotechnology), *CLEC4F* major (5'-TGTCTTGGTTTCTGTTTCCCCATAC-3'; chr2:71,050,550-71,050,574), *CLEC4F* minor (5'-TGTCTTGGTTTATGTTTCCCCATAC-3'; chr2:71,050,550-71,050,574), *ARL5B* major (5'-GACATTCAGTTTCGTTTCATGCCAG-3'; chr10:18,947,371-18,947,395) and *ARL5B* minor (5'-GACATTCAATTTCGTTTCATGCCAG-3'; chr10:18,947,371-18,947,395). 1 ug nuclear extract was incubated with 1 ng dsDNA probe for 30 min at 21°C in buffer consisting of 10 mM Tris-HCl (pH 7.5), 50 mM NaCl, 1 mM DTT, 1 mM EDTA and 12.5% glycerol (Sigma-Aldrich). 1 ug antibody was added for super-shift reactions. Reactions were resolved by native PAGE (NativePAGE; Life Technologies) and autoradiographed.

CRISPR assay

CRISPR conversion of rs11080327 was performed as described (*69*). Briefly, a guide sequence (5'-CACCGCCCATGAGACAGACAGATT-3'; or 5'-CACCGAGACAGACA GATTAGGAAT-3' in independent line) directed to cleave the rs11080327 genomic locus was cloned into the Cas9-containing plasmid, PX330. The guide sequence-containing PX330 plasmid as well as a 99-base pair ssDNA oligo (5'-GAGAAGATGAGAGCAATGGCTAAAGTTTGTCTGGCCTCGCCCTATGAGAAGG GGGCGGAGGCTGTGCCTGCCTTTGTGGGGGAAAAAGAAgCCGATTCCTAATC TGTCTGTCTCATGGGAGGTGAGAGTTCCTGTGGCAGGATAACAGACAAGCAA TTCAGTGTGATAACATCCTGTGTCA-3') – used as the template for repair to edit rs11080327$^{A/G}$ to rs11080327$^{G/G}$ – were transfected into HEK293T cells using Lipofectamine LTX (Life Technologies). Cells were stimulated with 1000 U/mL recombinant IFNβ for 6.5 h, and then lysed in RLT buffer supplemented with 1% β-mercaptoethanol. Lysates were run on Nanostring.

Supplementary Text

Identification and population analysis of *cis*-eQTLs

As described in the main text, we identified 264 *cis*-eQTLs. To assess the sharing of eQTLs between conditions, we used a meta-analysis to find 173 genes with *cis*-eQTLs in resting, 215 in LPS-stimulated, 217 in influenza-stimulated and 200 in IFNβ-stimulated MoDCs (fig. S3C; *P*-value < $5x10^{-5}$; permutation FDR 0.04-0.05). In total, 29% of *cis*-eQTL associations were identified in stimulated but not in resting cells (M-value > 0.9).

Since the cohort consisted of three human populations, we assessed whether any *cis*-eQTLs were common or specific to these populations. First, to determine whether our *cis*-eQTLs replicated across populations, we used a meta-analysis with variants imputed for each population to find an average of 96% of *cis*-eQTLs (*P* < $1x10^{-7}$; permutation test, FDR < $4x10^{-4}$) detected in at least two populations (M-value > 0.9; table S5), suggesting a high level of replication within our study. Second, we identified 121 genes differentially expressed between populations in at least one condition, of which ~50% have significant *cis*-eQTLs (table S5). Since differences in gene expression between human populations are believed to be predominantly due to minor allele frequency (MAF) differences between populations (*70*), we identified the subset of genes that exhibit significant population differences in gene expression and MAF (P_{ST}; *P* < 0.05; F_{ST}, *P* < 0.05) (fig. S3F). Among the 10 identified genes, *IFITM3* was associated with a variant (rs7944394) found in much higher frequency in Caucasians (MAF 0.44 EUR, 0.1 AFR and 0 ASI; F_{ST} = 0.30; F test, *P* < 0.05) with the minor allele associated with reduced *IFITM3* expression (fig. S3G). Consistent with significant variation (F test, *P* < 0.05) in allele frequency across human populations, this anti-viral effector gene has been proposed to be subject to positive selection (*71, 72*).

Most genes with *cis*-reQTLs have baseline eQTLs

We note that since 81 of the 121 genes with *cis*-reQTLs also had significant *cis*-eQTLs in baseline (table S4), most of the *cis*-reQTLs do not arise from an increase in expression or power in stimulation conditions. In agreement, we did not detect a strong bias in our ability to detect *cis*-reQTLs based on the magnitude of absolute expression or differential expression (fig. S3H). Thus, reQTLs like arise from gene-by-environment interactions.

Identification of multiple *cis*-eQTLs in *TEC* locus

Expression variation of the protein tyrosine kinase gene, *TEC*, illustrates the regulatory complexity of these loci. In the *TEC* locus, we detected the 3' SNP rs10938526 and the 5' SNP rs2271173 (conditioned on rs10938526) as two independent *cis*-eQTLs with additive effects (fig. S3D): the variant at the 3' end (rs2271173) is a *cis*-eQTL associated with baseline expression while the variant at the 5' end (rs10938526) is a *cis*-eQTL associated with *TEC* expression in the LPS and influenza conditions but not in the baseline nor IFNβ conditions (fig. S3B). Thus, the presence of multiple *cis*-eQTLs per gene reflects regulatory complexity and helps explain pathway-specific variation in gene expression.

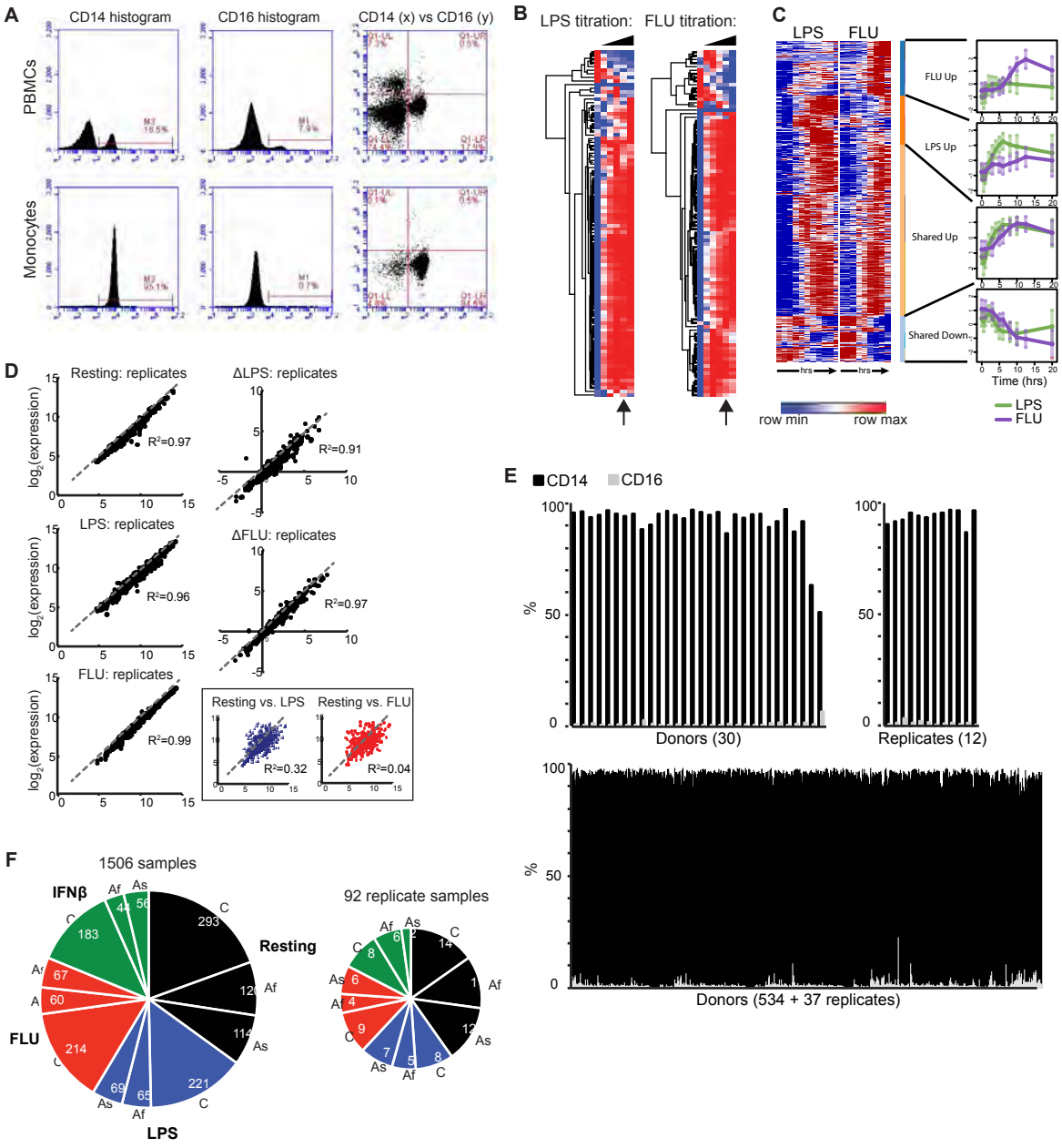


Fig. S1. A strategy to identify gene-by-environment interactions in the innate immune responses of primary human dendritic cells. (**A**) Anti-CD14 and anti-CD16 FACS plots showing purity of purified $CD14^{+}CD16^{lo}$ monocytes vs. PBMCs. (**B**) Heatmap of gene expression (log_2(nCounts)) of 96 genes (pilot Nanostring gene set) from MoDCs pooled from 5 donors and left unstimulated or stimulated with LPS or infected with influenza virus over a titration curve. Concentrations indicated with arrows were selected for subsequent assays. (**C**) Heatmap of microarray expression from MoDCs (pooled from 13 donors) stimulated over a detailed timecourse with FLU (0, 1, 3, 5, 6, 8, 10, 12.5 and 20 h) or LPS (0, 0.5, 1, 2, 3, 4, 5, 6, 8, 10 and 20 h). Right: mean expression (y-axis; LPS, green; FLU, purple) and standard deviation (error bar) at each time point (x-axis) of FLU-upregulated, LPS-upregulated, LPS- and FLU-upregulated (shared), and

LPS- and FLU-downregulated clusters. Time points where FLU response (10 h) and LPS response (5 h) plateau are highlighted and were used for subsequent assays. (**D**) Expression (log_2(expression)) of 415 reporter genes in MoDCs prepared from two different vials of the same PBMC source to estimate the technical reproducibility of the assay. Monocytes from the two samples were separately isolated, differentiated into MoDCs, and left unstimulated or stimulated with LPS or FLU. Gene expression was measured by microarray, expression or induction of the two samples was plotted, and R^2 was calculated. For comparison, expression of resting MoDCs vs. of LPS-stimulated MoDCs is shown in inset in blue, and expression of resting MoDCs vs. of FLU-infected MoDCs is shown in inset in red. (**E**) Anti-CD14 and anti-CD16 FACS data from the initial 30 donors and 12 replicate samples (top) for the microarray study, and from the subsequent 534 donors and 37 replicate samples (bottom) for the Nanostring study. (**F**) Pie charts showing the 1506 samples as well as 92 replicate samples (total 1598 samples) separated by condition and ethnicity of individuals; black, resting; blue, LPS; red, FLU; green, IFNβ; C, Caucasian; Af, African-American; As, Asian.

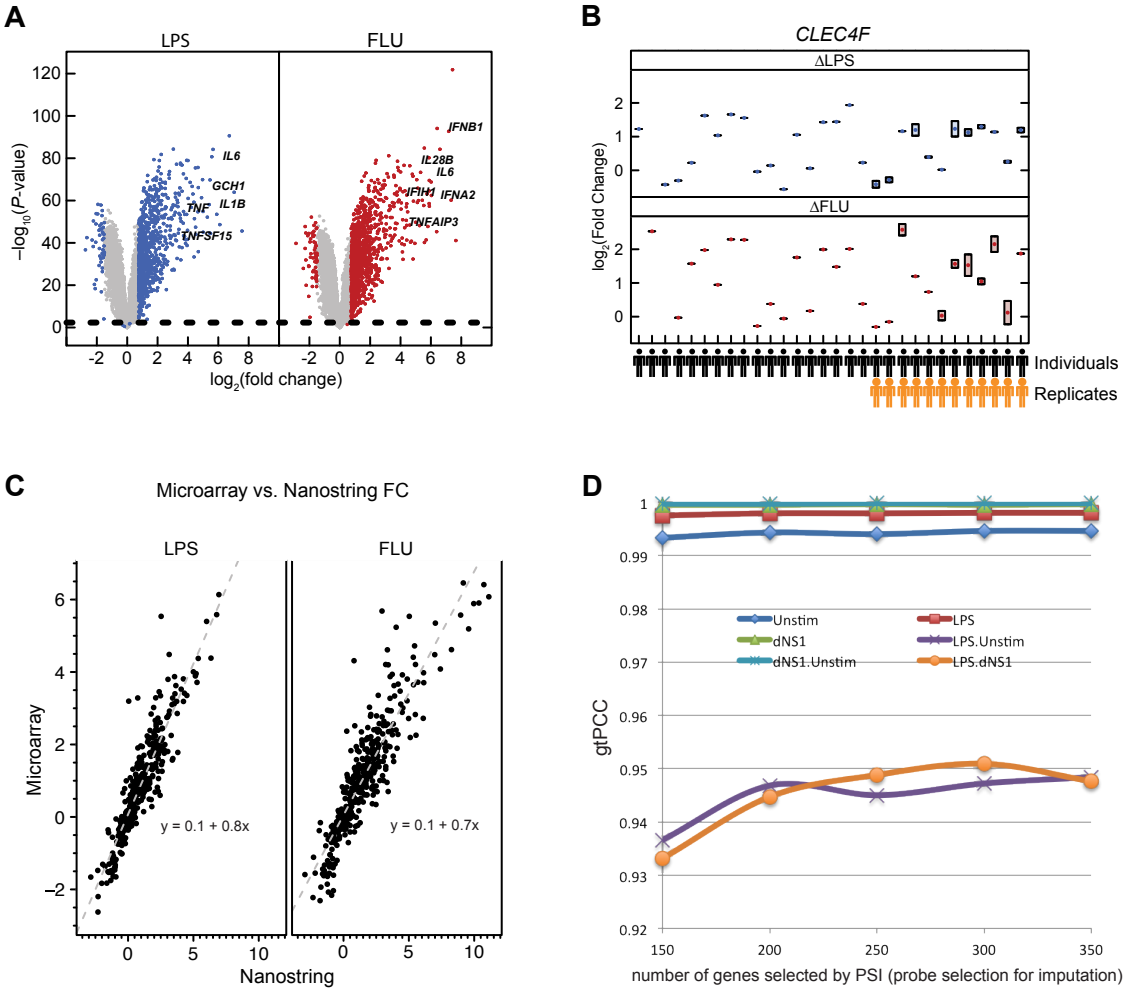
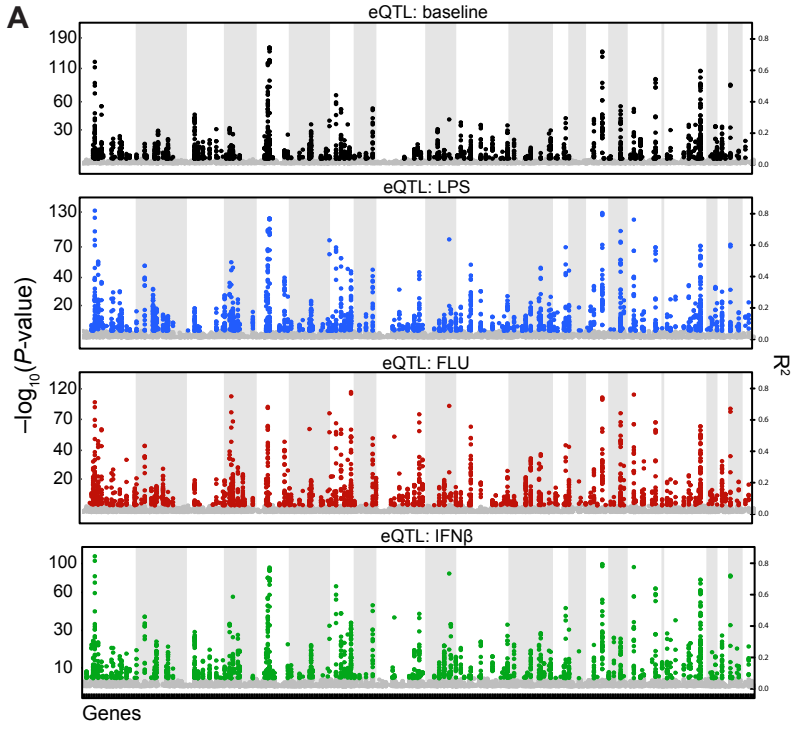
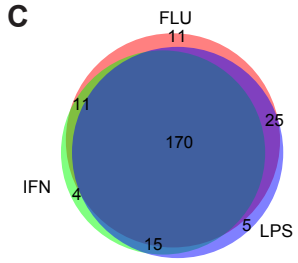
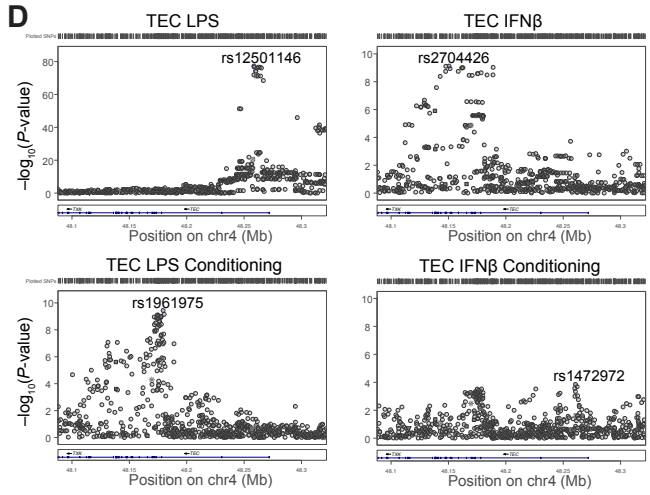
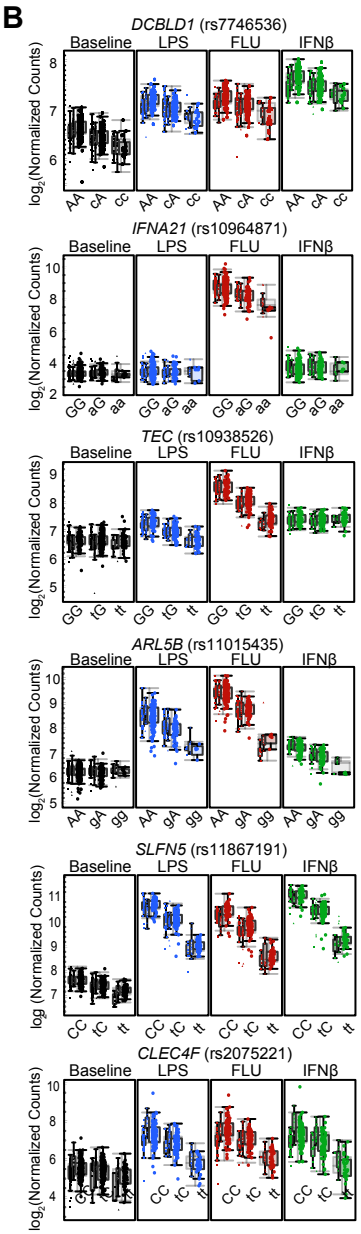


Fig. S2. Genome-wide expression profiles in MoDCs reveal response phenotypes. (**A**) Volcano plots showing differentially expressed genes in MoDCs, from 30 individuals, stimulated with LPS for 5 h or FLU for 10 h. The mean $\log_2$(fold change, stimulated/unstimulated) in expression of each gene (x-axis) is plotted against the $-\log_{10}$(*P*-value) testing significance of fold change (y-axis). Genes showing $\log_2$(fold change) greater than 0.75 or less than –1.5 are shown as colored dots (blue, LPS-stimulated; red, FLU-stimulated). Selected genes are labeled. (**B**) Log_2(fold change) of *CLEC4F* in LPS-stimulated and FLU-infected MoDCs from 30 donors and 12 replicates. Standard error of replicate samples is shown for each sample. Fold change of *CLEC4F* shows significant (FDR < 0.01) inter- vs. intra-individual variation following LPS and FLU stimulations. (**C**) LPS-induced (left) and FLU-induced (right) fold changes of Nanostring reporter genes are plotted against fold changes of respective genes from microarray. (**D**) Ground truth Pearson correlation coefficient (gtPCC) (y-axis) plotted against number of genes selected for codeset (x-axis), showing that most of the variance in the expression data is captured with > 150 probes.


A
eQTL: baseline
eQTL: LPS
eQTL: FLU
eQTL: IFNβ
−log10(P-value)
R2
Genes


C
FLU
11
11
25
170
IFN
4
5
LPS
15


D
TEC LPS
rs12501146
TEC IFNβ
rs2704426
TEC LPS Conditioning
rs1961975
TEC IFNβ Conditioning
rs1472972
−log10(P-value)
Position on chr4 (Mb)


B
DCBLD1 (rs7746536)
IFNA21 (rs10964871)
TEC (rs10938526)
ARL5B (rs11015435)
SLFN5 (rs11867191)
CLEC4F (rs2075221)
Baseline
LPS
FLU
IFNβ
log2(Normalized Counts)

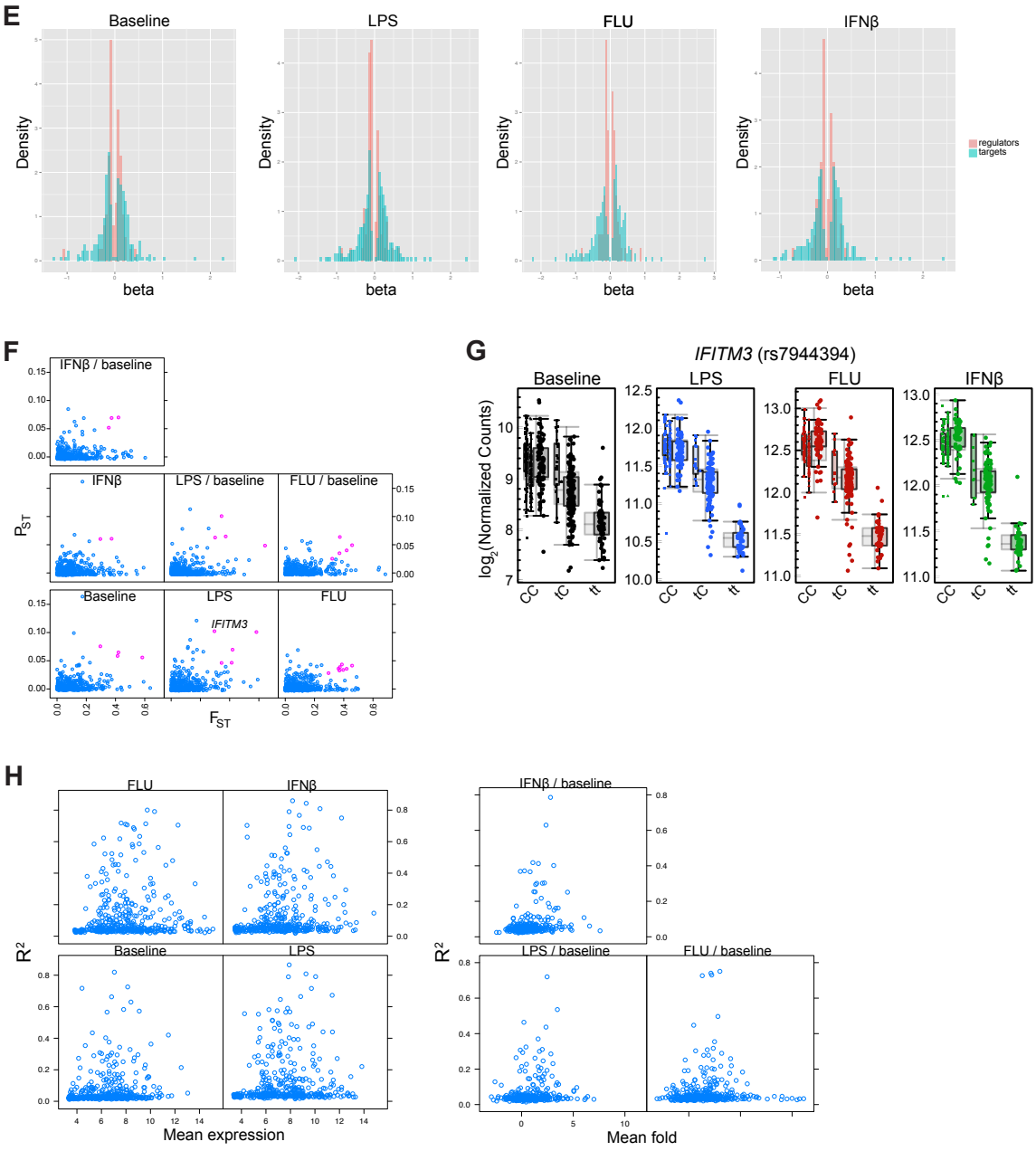


Fig. S3. Association analysis reveals *cis*-eQTLs and *cis*-reQTLs. (**A**) Manhattan plot of *cis*-eQTLs in baseline, LPS-, FLU- and IFNβ-stimulated MoDCs, showing $-\log_{10}$(*P*-values) (left y-axis) and R^2 values (right y-axis) for all *cis*-SNPs, displayed on the x-axis with associated genes ordered by chromosomal location. (**B**) Box-whisker plots showing expression ($\log_2$(nCounts)) of *DCBLD1*, *IFNA21*, *TEC*, *ARL5B*, *SLFN5* and *CLEC4F* in resting, LPS-stimulated, FLU-infected and IFNβ-stimulated MoDCs as a function of genotype of the respective *cis*-SNPs (x-axis: rs27434, rs10964871, rs10938526, rs11015435, rs11867191 and rs2075221). African-Americans, Asians, and Europeans in this order are displayed as separate box-whisker plots adjacent to each other in each condition. (**C**) Venn diagrams showing overlap of *cis*-eQTLs between stimulation conditions determined by meta-analysis and M-value (M-value > 0.9 used as inclusion criteria). (**D**) LocusZoom plots showing the $-\log_{10}$(*P*-values) of imputed *cis*-eQTL (y

axis) in the chromosomal regions (x-axis) of *TEC* before (top) and after (bottom) conditioning on the five most significant SNPs in each region. Independent signals were seen in in *TEC* after LPS but not IFNβ stimulation. (**E**) Histogram of *cis*-eQTL effect sizes of regulators (red) and targets (blue), showing that effect sizes of regulators are lower than effect sizes of targets. (**F**) P_{ST} plotted against F_{ST}, highlighting genes with significant expression differences between populations. Purple dots represent genes with significant ($P < 0.05$) P_{ST} and F_{ST} values. (**G**) Expression ($\log_2$(nCounts); y-axis) of *IFITM3* in resting, LPS-stimulated, FLU-infected and IFNβ-stimulated MoDCs as a function of genotype of rs7944394, showing that *IFITM3* exhibits a population-specific *cis*-eQTL due to different MAFs between populations. African-Americans, Asians, and Europeans are displayed as separate box-whisker plots adjacent to each other in each condition. (**H**) R^2 as a function of expression (left) and differential expression (right), showing the lack of expression bias of eQTL and reQTL signals.

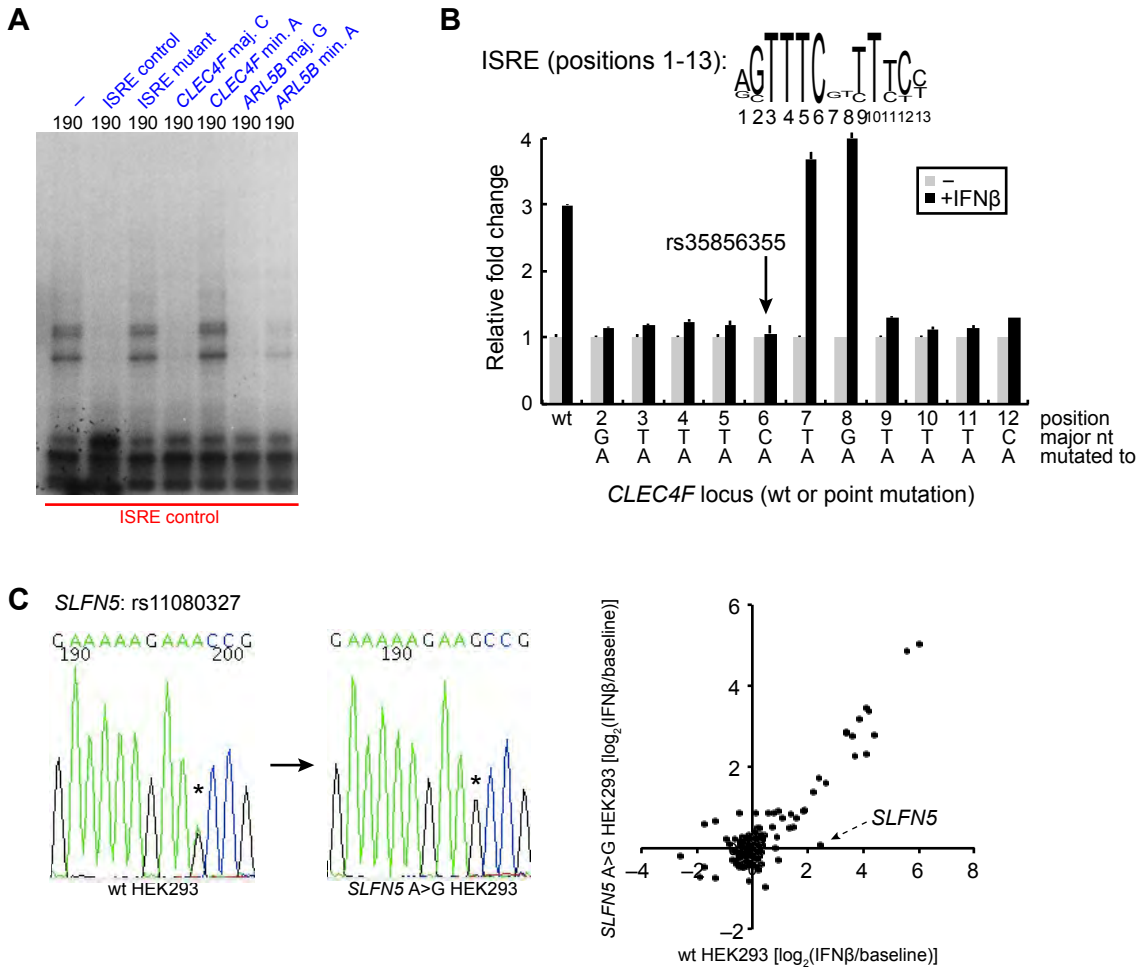


Fig. S4. Functional fine-mapping and mechanism of *cis*-reQTLs. (**A**) Electrophoretic mobility shift assays (EMSA) with 27 bp radiolabelled dsDNA probe containing a known ISRE motif control incubated with nuclear lysates from IFNβ-stimulated MoDCs. Competition assays were performed with 100-fold excess of cold probes corresponding to the ISRE motif control, a mutated ISRE motif control, the *CLEC4F* rs35856355 major (G) or minor allele (T) sequences, or the *ARL5B* rs2130531 major (G) or minor allele (A) sequences. (**B**) Luciferase expression from reporter constructs transfected into HEK-293 cells that were left unstimulated or were stimulated with 1000 U/mL IFNβ for 21 h. 159 bp sequence from wt *CLEC4F* reporter construct, or *CLEC4F* reporter constructs with single point mutations along ISRE motif, were subcloned 5' of a minimal promoter and firefly luciferase gene. Firefly luciferase expression was normalized to *Renilla* luciferase expression expressed from co-transfected plasmid. Mean and s.d. are graphed. (**C**) 5.8 kb (chr17:33,568,021-33,573,886) surrounding rs11080327 – a region containing an H3K27Ac signal from ENCODE – was sequenced in both wild type and CRISPR-converted lines, and no differences except for rs11080327 were detected. To further confirm the results, an alternative guide sequence was used to create an independent CRISPR-converted line: fold change $\log_2$(IFNβ stim / unstim) of signature genes in wild type HEK-293 cells (rs11080327$^{A/G}$) is plotted against fold change in CRISPR-converted (rs11080327$^{G/G}$) HEK-293 cells created using the alternative guide sequence.

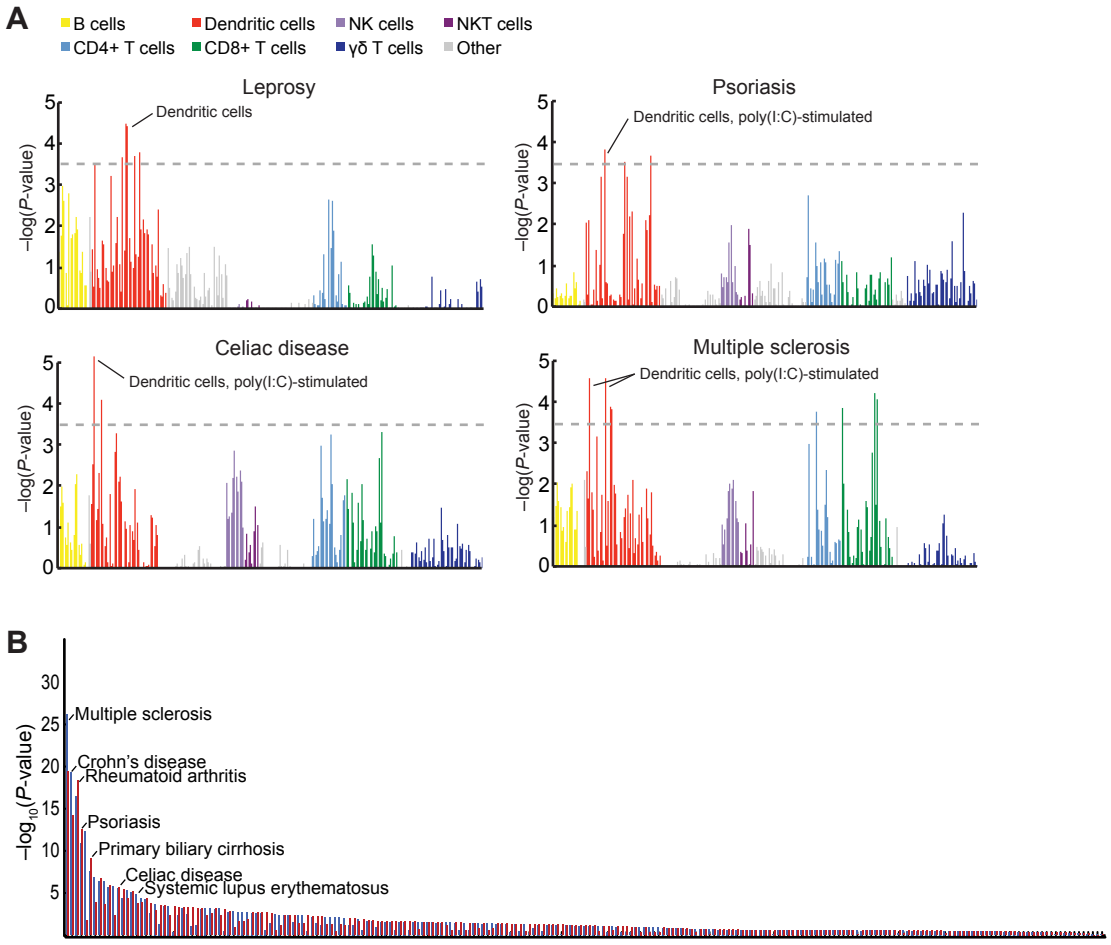


Fig. S5. Autoimmune and infectious disease-associated SNPs are *cis*-eQTLs and *cis*-reQTLs. (**A**) Evidence for enrichment of cell-type-specific expression of genes near GWAS hits in resting and stimulated dendritic cells, relative to 248 immune cell types in the ImmGen database. Each bar represents the empirical *P*-value for a single cell type, colored by cell type groupings. The dashed line is a Bonferroni-corrected significance threshold for the number of cell types tested. (**B**) Ranking of GWAS diseases by the hypergeometric *P*-value of enrichment for variants in close proximity to induced genes ($\log_2$(fold change) > 0.75) in LPS stimulation (blue) and FLU infection (red).

Table S1. Donors and samples used in study. Description of samples used in microarray and Nanostring experiments including demographic information of donors, cell counts, cell purity, annotation of serial replicates, and stimulation conditions.

Table S2. Differential expression analysis of microarray data after LPS and FLU stimulation. Differential expression results from analysis using Empirical Bayes (limma package) is reported with –log(fold change), –log(*P*-value) and –log(FDR) after LPS or FLU stimulation.

Table S3. Gene signature set. List and annotation of 415-gene signature set used for Nanostring gene expression profiles, including gene names, probe sequences, sources from how they were selected, FDR values for inter-individual and inter-population variability, and mean expression values.

Table S4. *cis*-eQTLs and *cis*-reQTLs from pooling three populations together using only tagging SNPs. The best tag *cis* SNP associated with the absolute expression or differential expression for each gene in each condition is reported, including effect size estimates, coefficient of determination, T statistic, point-wise *P*-value/FDR and permutation *P*-value/FDR. Also included is a meta-analysis on the absolute expression or differential expression over all conditions to determine effects that are shared across stimuli.

Table S5. Gene-wise *cis*-eQTLs from meta-analysis across populations after imputation. The best meta-analysis associated *cis*-eQTL for the absolute or differential expression of each gene in each condition is reported with fixed model and random model estimates for effect size, standard error, I squared, Cochran Q, meta-analysis *P*-value, population specific *P*-values and M-values.

Table S6. Conditioning analysis of *cis*-eQTLs from pooling three populations together using only tagging SNPs. The second best tag *cis* SNP associated with the absolute expression or differential expression for each gene in each condition is reported including effect size estimates, coefficient of determination, T statistic, point-wise *P*-value/FDR and permutation *P*-value/FDR.

Table S7. ChIP-Seq enrichment of top 5 associated imputed SNPs for each gene. Enrichment calculated as fold change over background and the respective binomial *P*-value for each condition.

Table S8. *Trans*-eQTLs and *trans*-reQTLs from pooled analysis (all variants). The best meta-analysis associated *trans*-eQTL for the absolute or differential expression of each gene in each condition is reported with fixed model and random model estimates for effect size, standard error, I squared, Cochran Q, meta-analysis *P*-value, population specific *P*-values and M-values.

Table S9. *Trans*-eQTLs and *trans*-reQTLs from pooled analysis (only *cis* variants considered). The best meta-analysis associated *trans*-eQTL, when considering only the set of SNPs *cis* to at least one gene on our gene signature set, for the absolute or differential expression of each gene in each condition is reported with fixed model and random model estimates for effect size, standard error, I squared, Cochran Q, meta-analysis *P*-value, population specific *P*-values and M-values.

Table S10. *Trans* association of rs12805435. Association statistics of rs12805435 to gene expression in each condition.

Table S11. GWAS enrichment. Overlap of GWAS variants with *cis*-eQTLs, with significant diseases highlighted determined by hypergeometric score.

References

46. S. D. Shapira *et al.*, A physical and regulatory map of host-influenza interactions reveals pathways in H1N1 infection. *Cell* **139**, 1255 (Dec 24, 2009).
47. B. P. Fairfax *et al.*, Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. *Nature genetics* **44**, 502 (May, 2012).
48. L. B. Barreiro *et al.*, Deciphering the genetic architecture of variation in the immune response to Mycobacterium tuberculosis infection. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 1204 (Jan 24, 2012).
49. T. Kawai, S. Akira, The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. *Nature immunology* **11**, 373 (May, 2010).
50. O. Takeuchi, S. Akira, Pattern recognition receptors and inflammation. *Cell* **140**, 805 (Mar 19, 2010).
51. R. Barbalat, S. E. Ewald, M. L. Mouchess, G. M. Barton, Nucleic acid recognition by the innate immune system. *Annu Rev Immunol* **29**, 185 (2011).
52. G. R. Stark, I. M. Kerr, B. R. Williams, R. H. Silverman, R. D. Schreiber, How cells respond to interferons. *Annual review of biochemistry* **67**, 227 (1998).
53. Y. Donner, T. Feng, C. Benoist, D. Koller, Imputing gene expression from selectively reduced probe sets. *Nature methods* **9**, 1120 (Nov, 2012).
54. G. K. Geiss *et al.*, Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nature biotechnology* **26**, 317 (Mar, 2008).
55. S. Anders, W. Huber, Differential expression analysis for sequence count data. *Genome biology* **11**, R106 (2010).
56. A. L. Price *et al.*, Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics* **38**, 904 (Aug, 2006).
57. S. Purcell *et al.*, PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics* **81**, 559 (Sep, 2007).
58. A. A. Shabalin, Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* **28**, 1353 (May 15, 2012).

59. J. D. Storey, R. Tibshirani, Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 9440 (Aug 5, 2003).
60. B. Han, E. Eskin, Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *American journal of human genetics* **88**, 586 (May 13, 2011).
61. B. L. Browning, S. R. Browning, A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *American journal of human genetics* **84**, 210 (Feb, 2009).
62. L. C. Tsoi *et al.*, Identification of 15 new psoriasis susceptibility loci highlights the role of innate immunity. *Nature genetics* **44**, 1341 (Dec, 2012).
63. N. Zaitlen, B. Pasaniuc, T. Gur, E. Ziv, E. Halperin, Leveraging genetic variability across populations for the identification of causal variants. *American journal of human genetics* **86**, 23 (Jan, 2010).
64. E. P. Consortium *et al.*, An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57 (Sep 6, 2012).
65. L. D. Ward, M. Kellis, HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic acids research* **40**, D930 (Jan, 2012).
66. M. Garber *et al.*, A high-throughput chromatin immunoprecipitation approach reveals principles of dynamic gene regulation in mammals. *Molecular cell* **47**, 810 (Sep 14, 2012).
67. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841 (Mar 15, 2010).
68. C. Genomes Project *et al.*, An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56 (Nov 1, 2012).
69. L. Cong *et al.*, Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819 (Feb 15, 2013).
70. B. E. Stranger *et al.*, Patterns of cis regulatory variation in diverse human populations. *PLoS genetics* **8**, e1002639 (2012).
71. A. L. Brass *et al.*, The IFITM proteins mediate cellular resistance to influenza A H1N1 virus, West Nile virus, and dengue virus. *Cell* **139**, 1243 (Dec 24, 2009).
72. A. R. Everitt *et al.*, IFITM3 restricts the morbidity and mortality associated with influenza. *Nature* **484**, 519 (Apr 26, 2012).